Final author draft only. The copyedited article may differ from this manuscript version. The details of the article are as follows:

Litchfield D. & Donovan, T. (in press). Expecting the initial glimpse: prior target knowledge activation or repeated search does not eliminate scene preview search benefits. *Journal of Cognitive Psychology*.

Running Head: Scene-target and target-scene processing

Expecting the initial glimpse: prior target knowledge activation or repeated search does not eliminate scene preview search benefits

Damien Litchfield[1] *

Tim Donovan[2]

[1]Department of Psychology

Edge Hill University, UK

[2]Medical & Sport Sciences,

University of Cumbria, UK

*Corresponding author:

Department of Psychology, Edge Hill University, Ormskirk, L39 4QP

Email: damien.litchfield@edgehill.ac.uk

Tel: +44 (0)1695 584085

Abstract

A brief glimpse of a scene can guide subsequent eye movement behaviour but it is still unclear how prior activation of target knowledge influences early scene processing and later eye movements. Using the gaze-contingent 'flash-preview moving window' (FPMW) paradigm to restrict access of peripheral vision during visual search, we manipulated whether identity of search targets was presented before or after scene previews. As expected, windowed search for a target was more efficient following a 250ms scene preview, whereas knowing the target identity prior to scene preview led to further improvements in how search was initiated and executed. However, in Experiment 2 when the target was not present during scene previews, only metrics reflecting the initiation of search continued to be modulated by prior activation of target knowledge. Experiment 3 provided new evidence to suggest that the search benefits from scene previews are maintained even when participants are repeatedly searching through the same type of scene for the same type of target. Experiment 4 replicated Experiment 3 whilst also controlling for differences in integration times. We discuss the robust findings of scene preview search benefits and the flexibility of the FPMW paradigm to measure how the first glimpse affects search.

Keywords: flash-preview moving window; eye movements; scene perception; expectation; visual search

Introduction

We live in a rich visual world with a multitude of items competing for our attention. How we decide where to look depends on our experience, our current task-goals and how we first perceive our environment (Buswell, 1935; Henderson, 2007; Rayner, Smith, Malcom, & Henderson, 2009; Yarbus, 1967). Within a brief glimpse of a real-world scene we can rapidly recognise the scene category, or ‘scene-gist’, and infer what objects would be in such scenes (Biederman, Mezzanotte, & Rabinowitz, 1982; Intraub, 1980) and where they are likely to be located (Castelhano & Henderson, 2007; Koehler, & Eckstein, 2017). In addition to rapid scene recognition, this initial representation of the scene can be retained in memory and used to plan subsequent eye movements (Castelhano & Henderson, 2007; Hollingworth, 2009; Wolfe, Võ, Evans, & Greene, 2011). Research using the ‘flash-preview moving window’ (FPMW) paradigm (Castelhano & Henderson, 2007; for an overview see also Litchfield & Donovan, 2017) has been instrumental in furthering our understanding of scene perception as it enables the initial scene representation to be dissociated from the ongoing scene representation obtained during search.

The FPMW is a gaze-contingent paradigm where observers are shown a brief preview of a scene (or control) and then asked to search for a target object within the same scene whilst their peripheral vision is restricted. Restricting peripheral vision during search ensures observers must rely on what they could process during the initial scene preview to further guide search, otherwise it is difficult to isolate the specific contribution of the initial scene preview on subsequent eye movement behaviour. With scene previews, eye movement metrics indicate windowed search is more efficiently initiated and executed and that this is due to the initially generated scene representation interacting with the target knowledge that is activated from viewing the subsequently presented target word (Hillstrom, Schloley, Liversedge, Benson, 2012), or picture (Castelhano & Heaven, 2010). Moreover, these scene preview benefits predominantly rely on scene-context processing, rather than local target processing, because scene preview benefits are still found even when targets are digitally removed from previews and only found through windowed search (Castelhano & Henderson, 2007; Võ & Henderson, 2010).

The FPMW paradigm has revealed individual differences associated with initial scene processing (Litchfield & Donovan, 2016; Võ & Schneider, 2010), the extent to which semantically consistent objects are processed within scenes (Castelhano & Heaven 2011; Võ & Henderson, 2011), how learned object function may guide attention aside from object features (Castelhano & Witherspoon, 2016), and the time-course of the initial representation (Hillstrom, et al., 2012; Võ & Henderson, 2010). For example, Võ and Henderson (2010) found that although 250ms offered an optimal scene preview benefit, a shorter scene preview could guide search provided there was a corresponding increase in the time between onset of scene preview and the start of windowed search (i.e., integration time). Võ and Henderson (2010) suggested that this increase in integration time enables the activation of knowledge and expectations regarding scenes and the target to constrain search (Torralba, Oliva, Castelhano, & Henderson, 2006). Moreover, increased integration time reduces search times and also leads to stronger memory representations when compared to deliberate memorization tasks (Josephs, Draschkow, Wolfe, & Võ, 2016).

With one exception (Litchfield & Donovan, 2016), FPMW studies have thus far only studied the integration of scene and target knowledge post preview. That is, a scene preview is presented first and then target identity is revealed to participants after the preview, but prior to search. This approach has its merits in revealing the time-course of the initial representation, and how the visual system can rapidly process a random array of scenes and integrate the resulting representation with subsequently informed target information. However, manipulating target knowledge to be activated only after a scene is presented somewhat undermines the ever-present role of expectation, in that much of what we see is what we expect to see (Summerfield & Egner, 2009; Summerfield, & de Lange, 2014). Our goals to find a target or to perform a particular action often precede the initial scene representation, and these goals can help us quickly find what we are looking for (Hollingworth, 2009), but also blind us to what we do not expect to see (Drew, Võ, & Wolfe, 2013; Simons, & Chabris, 1999). Moreover, having the search target identity only revealed after the scene preview is at odds with the eye-tracking data used to benchmark highly successfully contextual guidance models (e.g., Torralba et al., 2006), or the LATEST model used to predict where and when people look in scenes (Tatler, Brockmole & Carpenter, 2017). This is because such eye-tracking data is from observers that either knew the target identity before the scenes were presented or knew that they had to remember the scenes.

Litchfield and Donovan (2016) were forced to address this issue of target identity being known before scene preview because they applied the FPMW paradigm to medical image perception, and in such tasks the observer typically knows beforehand what type of image they are likely to be presented with and what type of target (abnormality) they will be looking for (Donovan & Litchfield, 2013; Drew et al., 2013). Given that experts and novices have different levels of target/scene knowledge of medical images, the goal of their study was to use the FPMW to systematically isolate how this expertise-dependent knowledge contributed to initial scene processing, subsequent eye movement behaviour and decision-making performance. Litchfield and Donovan (2016) found that when searching for everyday objects from real-world scenes, experts in medical image perception were no different from psychology students when it came to exploiting scene previews to enhance search performance. When searching for domain-specific targets (lung cancer) from domain-specific images (chest x-rays) experts correctly identified more targets than novices but experts were not more accurate if given a scene preview. In terms of search benefits from scene preview, there was only a weak expertise advantage, with experts being slightly more efficient at searching for these targets under FPMW conditions. These recent findings are at odds with the dominant theory in medical image perception (Nodine & Kundel, 1987; Kundel, Nodine, Kunant, & Weinhaus, 2007) which places emphasis on experts being able to outperform novices in exploiting brief glimpses of medical images to rapidly detect abnormalities. It is important to determine whether our previous findings are not simply a result of methodological issues that occurred when adapting the FPMW paradigm. The purpose of the present study is to resolve three key FPMW methodological issues that have broad implications for real-world scene perception.

The first issue is whether target identity is known before or after scene previews. In the FPMW sequence used by Litchfield and Donovan (2016) target identity was always known before the scene previews for both the scene-perception and the medical task, thus providing a clear comparison between these two tasks. Robust scene preview effects were found for the scene-perception task, however, studies have yet to directly compare the scene preview effects in FPMW when target identity is known before vs. after scene preview.

The second issue relates to whether the target is visible or not during scene previews, and how this may interact given the first issue of knowing the target identity beforehand. It has been established that the scene preview benefit is the same even when the target is not visible during the initial glimpse (Castelhano & Henderson, 2007; Võ & Henderson, 2010), but in these studies the participants were not told the target identity until after the preview, thus requiring participants to always integrate such target information with a previously acquired visual representation. However, there are many circumstances where target identity is known before the scene is presented and this will change what observers expect to see and may influence how the initial scene is processed (Summerfield & Egner, 2009). For example, target detection processes can begin immediately within the first of glimpse of the scene, but not so with existing FPMW studies that only reveal target identity after the preview. If the target was visible during preview we would predict that this would lead to further improvements in search when coupled with prior activation of target knowledge. But what is not yet clear is if there is a difference in search when directly comparing target knowledge activation before or after preview, and the target is not visible during preview. Knowing what the target is beforehand may enable target detection processes to be activated within the glimpse but may also activate scene knowledge and give the observer an expectation of the upcoming scene. Therefore, even when the target is not visible, the observers may encode the glimpse differently, and potentially more efficiently, when they have some idea of what the scene is likely to be.

The third issue is that in the current implementations of the FPMW paradigm there is always a random array of scenes and targets that are unique to each trial, thereby maximising the benefit of seeing the scene preview before search. Although the scene preview benefit has so far been a robust effect that diminishes after the first few fixations (Hillstrom et al., 2012), it is not yet clear how the scene preview fares under repeated search conditions. Litchfield & Donovan (2016) tried to examine this issue of repeated search on scene previews in their applied medical image perception task, but task difficulty may have overshadowed their findings and therefore it remains unresolved whether repeated search diminishes the scene preview benefit when using real-world scenes. We therefore aim to establish whether there is still a scene-preview benefit when all FPMW trials involve searching through the same type of scene for the same type of target.

To examine these 3 research issues we devised 3 experiments that each make use of 4 within-participants conditions. Scene-Target and Mask-Target conditions represent the standard conditions used in the FPMW paradigm, whereby target identity is only revealed after a preview. As a comparison to these conditions, we also include Target-Scene and

Target-Mask conditions, whereby target identity is revealed before a preview. In Experiment 1, the target is visible during preview, whereas in Experiment 2, the target is digitally removed during preview. Experiment 3 investigates the issue of repeated search by deploying the same 4 conditions as Experiment 2, but instead of a random array of scenes and targets, participants repeatedly search through kitchen scenes for cups. Experiment 4 replicated the effects of Experiment 3 whilst also controlling for the differences in integration times that can arise when manipulating whether target identity is made known before or after preview.

Experiment 1

Method

Participants

Twenty Lancaster University students participated in the experiment for course credit. All participants had normal or corrected-to-normal vision. This sample size is comparable to previous FPMW studies that found robust scene preview effects (e.g., Castelhano & Henderson, 2007; Võ & Henderson, 2010).

Stimuli and Apparatus

The stimuli were 80 full-colour photographs of real-world scenes taken from the LabelMe database, a repository of copyright-free images (Russell, Torralba, Murphy, & Freeman, 2008). Half were indoor scenes (e.g., kitchens, offices, living rooms) and half were outdoor scenes (e.g., streets, parks, coastlines). All scenes were presented on a 19-in. CRT monitor (1024 x 768 pixels, 120 Hz). The scenes subtended 24.12° x 18.24° of visual angle when viewed from 57 cm and each scene contained a single unique target object with an average size of 3.59° x 3.60°. Across all images there was an equal probability of the target occurring on either the left or right side of the image. The mask used in all experiments was created in Adobe Photoshop and consisted of a random array of coloured pixels.

Eye movements were recorded using an EyeLink 1000 desktop eye tracker (SR Research, Canada) and calibration points of eye position were only accepted if they had an average resolution < 0.5° visual angle. Rectangular interest areas were created that best fit each target object. The key performance metrics were Response Time (RT), defined as the time from the onset of the windowed search screen until button press, and Accuracy, (% targets correctly identified). To assess the efficiency of search we examined the time until first target fixation (search latency) and the number of fixations until first target fixation. Finally, we examined the initial saccadic latency and initial saccadic amplitude of the first eye movement as these measures represent the first response relating to the rapid processing of the scene preview and the readiness to initiate search (Anderson, Donk, & Meeter, 2016; Hillstrom et al., 2012; Litchfield & Donovan, 2016; Võ & Henderson, 2010).

Procedure

Eye movements were calibrated using a 9 point calibration and validation. Participants were instructed that they would have up to 15 seconds to search for a target from a real-world scene under windowed viewing conditions, and that on some trials they may be shown a brief glimpse of the upcoming scene before commencing search. Target identity order was not made explicit. Figure 1 indicates the trial sequence for each condition. The key difference between conditions was whether the flash preview was the scene or a mask, and whether the target word was presented before or after the preview. When the target word was after preview (typical FPMW), participants were presented with a fixation cross for 200ms, and then shown either a scene preview or mask preview for 250ms, followed by a mask for 50ms. A black target word which indicated the identity of the target object was then presented for

1500ms, followed by a fixation cross for 400ms, after which, windowed search of the scene began. When the target word was shown before preview, participants were presented with a fixation cross for 200ms, and then shown the target word for 1500ms. Following a fixation cross for 200ms, either a scene preview or mask preview was presented for 250ms, followed by a mask for 50ms. A final fixation cross was presented for 400ms, after which, windowed search of the scene began. A 2.5° radius window was used to restrict the field of view during search and to detect a target participants pressed a gamepad button while directly fixating the target. Eight practice trials were presented to familiarise participants with the procedure followed by 4 blocks of 20 trials. Since targets were only considered correctly identified if a fixation was within the target AOI during button press, recalibrations took place between the 4 blocks of trials to ensure the window was correctly aligned throughout the experiment. Participants saw each scene once and scenes were rotated across participants via a Latin square. Conditions were randomly presented and the whole experiment took approximately 30min.

<< Insert Figure 1 about here >>

Results

All data were subjected to 2 x 2 repeated measures ANOVA with preview (scene, mask) and target order (before preview, after preview) as within-subjects factors. All comparisons for the interactions utilized the Bonferroni correction. For all measures only trials with correct responses and valid data[1] were analysed. With the exception of accuracy, we expected that all metrics would be more efficient with scene preview than mask preview, and that knowing the target identity before the flash preview would lead to further benefits in performance. A summary of means can be seen in Table 1.

<< Insert Table 1 about here >>

Performance

Search accuracy averaged 84% and did not differ as a function of preview $F(1, 19) = 3.25$, $p = .087$, $\eta_p^2 = 0.15$, target order, $F(1, 19) = 1.03$, $p = .32$, $\eta_p^2 = 0.05$, and there was no interaction, $F(1, 19) = 0.50$, $p = .83$, $\eta_p^2 < 0.01$. RTs averaged 3949ms across conditions and there was a main effect of preview, $F(1, 19) = 35.99$, $p < .001$, $\eta_p^2 = 0.65$, with faster RTs for scene preview than mask preview. However, there was no main effect of target order, $F(1, 19) = 1.19$, $p = .28$, $\eta_p^2 = 0.06$, and there was no interaction, $F(1, 19) = 1.86$, $p = .19$, $\eta_p^2 = 0.09$. Participants were faster to detect targets in scene preview than mask preview, however, target order did not affect RTs.

Search-related eye movements

Search latency averaged 2910ms across conditions and there was a main effect of preview, $F(1, 19) = 25.40$, $p < .001$, $\eta_p^2 = 0.57$, no main effect of target order, $F(1, 19) = 0.55$, $p = .46$, $\eta_p^2 = .02$, but there was a significant interaction, $F(1, 19) = 6.24$, $p = .022$, $\eta_p^2 = 0.25$. Search latencies to a target were faster for scene preview than mask preview

[1] In <1% of trials participants did not fixate the central fixation cross following a preview but instead were already fixating the target location (only found in target-scene and scene-target conditions). Since this would effectively negate many eye movement metrics, these trials were excluded from analysis. It should be noted therefore that some of the most powerful effects of target/scene guidance were not included in the analysis.

irrespective of whether the target word was presented before, $F(1, 19) = 27.91$, $p < .001$, $\eta_p^2 = 0.60$, or after preview, $F(1, 19) = 6.26$, $p = .022$, $\eta_p^2 = 0.25$. Moreover, as expected target order modulated search latencies with faster latencies for target-scene than scene-target conditions, $F(1, 19) = 5.71$, $p = .027$, $\eta_p^2 = 0.23$, but no such difference emerged between target-mask and mask-target conditions, $F(1, 19) = .60$, $p = .448$, $\eta_p^2 = 0.03$.

The number of fixations averaged 10.73 across conditions and there was a main effect of preview, $F(1, 19) = 34.52$, $p < .001$, $\eta_p^2 = 0.65$, and no main effect of target order, $F(1, 19) = 0.58$, $p = .45$, $\eta_p^2 = 0.03$. However, there was a significant interaction, $F(1, 19) = 8.01$, $p = .011$, $\eta_p^2 = 0.30$. Consistent with search latency, fewer fixations were needed for scene preview than mask preview irrespective of whether participants were presented with the target word before, $F(1, 19) = 37.16$, $p < .001$, $\eta_p^2 = 0.66$, or after a preview, $F(1, 19) = 6.92$, $p = .016$, $\eta_p^2 = 0.27$. In addition, presenting the target word before scene preview led to fewer fixations than presenting the target word after scene preview, $F(1, 19) = 7.13$, $p = .015$, $\eta_p^2 = 0.27$, but no such difference between such difference was found between target-mask and mask-target conditions, $F(1, 19) = .94$, $p = .344$, $\eta_p^2 = 0.05$.

First eye movement of search

The initial saccadic latency averaged 228ms across conditions and there was a main effect of preview, $F(1, 19) = 34.54$, $p < .001$, $\eta_p^2 = 0.65$. The main effect of target order was not significant, $F(1, 19) = 4.24$, $p = .053$. $\eta_p^2 = 0.18$, and there was no interaction, $F(1, 19) = 1.57$, $p = .23$, $\eta_p^2 = 0.08$. The first eye movement of search was executed much quicker in scene preview scene than mask preview.

The initial saccadic amplitude averaged 2.08° visual angle across conditions and there was a main effect of preview, $F(1, 19) = 42.58$, $p < .001$, $\eta_p^2 = 0.69$. There was also a main effect of target order, $F(1, 19) = 9.08$, $p = .007$, $\eta_p^2 = 0.32$, and there was also a significant interaction, $F(1, 19) = 17.35$, $p < .001$, $\eta_p^2 = 0.48$. Consistent with search latency and fixation frequency analyses, the first eye movement of search was much larger in amplitude in scene preview than mask preview, irrespective of whether participants were presented with the target word before, $F(1, 19) = 38.29$, $p < .001$, $\eta_p^2 = 0.66$, or after a preview, $F(1, 19) = 14.42$, $p < .001$, $\eta_p^2 = 0.43$. However, presenting the target word before scene preview led to a much larger first eye movement than presenting the target word after scene preview, $F(1, 19) = 19.80$, $p < .001$, $\eta_p^2 = 0.51$, but this was not the case when the preview was a mask, $F(1, 19) = 2.09$, $p = .164$, $\eta_p^2 = 0.10$.

By taking into account the direction and distance of the initial eye movement in relation to the subsequent distance to the target, we can also reveal whether these larger initial eye movements were actually honing in on the target location, or reflects a general readiness to fixate a potential target site in the periphery (i.e., a large, but misguided eye movement). The distance (in visual degrees) to the target was shortened as a function of preview, $F(1, 19) = 8.67$, $p = .008$, $\eta_p^2 = 0.31$, and target order, $F(1, 19) = 6.09$, $p = .023$. $\eta_p^2 = 0.24$, but there was no interaction $F(1, 19) = 3.21$, $p = .089$, $\eta_p^2 = 0.15$. Thus, when the target was visible during scene preview or known beforehand there appeared to be an advantageous shift towards the target within the first eye movement of search in Experiment 1.

Discussion

In line with previous research (e.g., Castelhano & Henderson, 2007; Litchfield & Donovan, 2016; Võ & Henderson, 2010), all metrics indicated search was more efficient when participants were presented with a scene preview than a mask preview. Following a 250ms glimpse of the upcoming scene, participants were quicker to initiate search and quicker to fixate the target. In addition, knowing target identity before scene preview led to further improvements in how search was executed, but such improvements did not carry over to RTs, suggesting this metric may not be as sensitive as search latency, as it also includes variability in verification time (Castelhano & Heaven, 2010; Castelhano & Witherspoon, 2016).

For mask-target and target-mask conditions, the first fixation of windowed search provides the first visual input of an otherwise shrouded scene. Consequently, it may be prudent to take further time extracting information from this restricted view of the scene in order to decide where to first move the window. This pattern of behaviour is reflected in the increase in initial saccadic latency and decrease in initial saccadic amplitude for mask preview relative to scene preview conditions, and has been observed elsewhere (e.g., Võ & Henderson, 2010). However, only initial saccadic amplitude was sensitive to target order manipulations, as the interaction showed that larger amplitudes followed scene previews, but a much larger first eye movement was made when the target was known before, rather than after scene preview.

Since target-features can be used as attentional guidance (e.g., Treisman, 2006), by keeping the targets visible within scene previews we anticipated that prior knowledge of the target would have helped constrain initial scene processing and enable target detection to occur based on the initial encoding of the scene preview. Consistent with this view, we found that the target-scene condition led to a larger first eye movement coupled with greater search efficiency compared to the scene-target condition. However, it remains to be the seen whether prior activation of target knowledge continues to influence how search is initiated and executed when target objects are not present within scene previews.

Experiment 2

Experiment 1 found that whether using the more traditional FPMW sequence, where target identity is known after preview, or the modified sequence where target identity is known before the preview, both result in scene preview effects. However, knowing the target identity before preview led to further improvements in initiating and executing search. Experiment 2 examined the relationship between target order and scene preview when targets were not visible during scene preview, and how this was reflected in how search was initiated and executed. Knowing the identity of the target prior to preview could have changed how the initial scene was encoded. However, if early object-detection was the sole reason for the additional improvements in the target-scene condition we should find no differences in performance between target-scene and scene-target conditions in Experiment 2.

Method

Participants

Twenty Lancaster University students participated in the experiment for course credit. All participants had normal or corrected-to-normal vision and did not participate in the previous experiment.

Stimuli

The stimuli and presentation was identical to Experiment 1 images with the exception that target objects were digitally removed from the scene previews using Adobe Photoshop.

Results

All data were subjected to 2 x 2 repeated measures ANOVA with preview (scene, mask) and target order (before preview, after preview) as within-subjects factors. For all measures, only eye movement metrics associated with correct responses were analysed. A summary of means can be seen in Table 2.

<< Insert Table 2 about here >>

Performance

Search accuracy averaged 84% and accuracy did not differ as a function of preview, $F(1, 19) = 2.27$, $p = .15$, $\eta_p^2 = 0.11$, target order $F(1, 19) = 0.12$, $p = .73$, $\eta_p^2 < 0.01$, and there was no interaction, $F(1, 19) = 0.13$, $p = .32$, $\eta_p^2 < 0.01$. RTs averaged 4021ms across conditions and there was a main effect of preview, $F(1, 19) = 29.30$, $p < .001$, $\eta_p^2 = 0.61$, a main effect of target order, $F(1, 19) = 7.23$, $p = .015$, $\eta_p^2 = 0.28$, but there was no interaction, $F(1, 19) = 1.12$, $p = .30$, $\eta_p^2 = 0.06$. Participants were faster to detect targets in scene preview than mask preview, and contrary to Experiment 1, RTs were faster when the target word was presented before a scene or mask preview.

Search-related eye movements

Search latency averaged 2930ms across conditions and there was a main effect of preview, $F(1, 19) = 17.20$, $p < .001$, $\eta_p^2 = 0.48$, however, there was no main effect of target order, $F(1, 19) = 2.83$, $p = .11$, $\eta_p^2 = .13$, and no interaction, $F(1, 19) = 0.98$, $p = .35$, $\eta_p^2 = 0.05$. Participants were faster to fixate targets in scene preview than mask preview. The number of fixations averaged 10.82 across conditions and there was a main effect of preview, $F(1, 19) = 15.26$, $p < .001$, $\eta_p^2 = 0.45$, but there was no main effect of target order, $F(1, 19) = 1.57$, $p = .23$, $\eta_p^2 = 0.08$, and no interaction, $F(1, 19) = 1.07$, $p = .31$, $\eta_p^2 = 0.05$. Targets were fixated in fewer fixations when participants were presented with a scene preview than a mask preview.

First eye movement of search

The initial saccadic latency averaged 242ms across conditions and there was a main effect of preview, $F(1, 19) = 44.05$, $p < .001$, $\eta_p^2 = 0.70$, a main effect of target order, $F(1,$

19) = 23.41, p = .004. η_p^2 = 0.52, but no interaction, $F(1, 19)$ = 2.40, p = .14, η_p^2 = 0.11. The first eye movement of search was executed 67ms quicker when the participants were presented with a scene preview than a mask preview. In addition, regardless of whether the preview was a scene or a mask, the first eye movement was 25ms quicker when the target word was presented before rather than after preview.

The initial saccadic amplitude averaged 2.19° visual angle across conditions and there was a main effect of preview, $F(1, 19)$ = 26.62, p < .002, η_p^2 = 0.58, a main effect of target order, $F(1, 19)$ = 7.54, p = .013, η_p^2 = 0.28, and a significant interaction, $F(1, 19)$ = 8.77, p = .008, η_p^2 = 0.32. The first eye movement of search was larger in scene preview than mask preview, irrespective of whether the target word was presented before, $F(1, 19)$ = 38.29, p < .001, η_p^2 = 0.66, or after preview, $F(1, 19)$ = 4.45, p = .048, η_p^2 = 0.19. Moreover, when the preview was the upcoming scene, initial saccadic amplitude was larger when the target word was presented before rather than after preview, $F(1, 19)$ = 12.51, p = .002, η_p^2 = 0.40, but this was not the case when the preview was a mask, $F(1, 19)$ = .96, p = .340, η_p^2 = 0.05. Unlike Experiment 1 these larger initial eye movements did not appear to be closing in on the direction of the target as distance to target was not shortened as a function of preview, $F(1, 19)$ = 3.45, p = .079, η_p^2 = 0.15, target order, $F(1, 19)$ = .56, p = .465. η_p^2 = 0.03, and there was no interaction $F(1, 19)$ = .03, p = .875, η_p^2 < 0.01.

Discussion

Removing the target objects from scene previews did not eliminate the scene preview benefit; search was still more efficient in scene preview than mask preview conditions and this supports previous findings on how scene-context alone can guide attention (Castelhano & Henderson, 2007; Võ & Henderson, 2010). However, the effects of target order on a subset of search efficiency metrics were diminished when targets were not visible as the selected eye movement metrics reflecting search efficiency showed no advantage of the target-scene condition over the scene-target condition. Thus, the success of the target-scene condition in Experiment 1 can be attributed to the detection of expected target-features present in the preview. Unlike Experiment 1, RTs were generally faster when the target word was presented before preview rather than after preview. Although this finding was not predicted, it again suggests that RTs are not simply a slower version of search latencies, as RTs also include the variability in verification time from first target fixation to pressing the response button (cf. Castelhano & Heaven, 2010).

Interestingly, the first of eye movement of search was still robustly modulated by target order as the first eye movement was faster and of larger amplitude when target identity was known before scene preview, even though the target was not present during preview. Our results show that when the target was visible during scene preview or known beforehand there appeared to be an advantageous shift towards the target within the first eye movement of search in Experiment 1. However, since no such effects were found for Experiment 2, the larger initial saccadic amplitude for the target-scene condition in Experiment 2 is more likely to reflect a general readiness to fixate a potential target site in the periphery, rather than an advantageous shift towards the target.

Experiment 3

Experiment 1 and 2 established that knowing the target identity before preview led to further improvements in initiating and executing search, but such effects were primarily exploited when the target was visible during the scene preview. However, typical with all current implementations of the FPMW paradigm, the previous experiments required participants to search for unique targets from unique scenes across all trials. Experiment 3 therefore establishes whether the scene preview benefit exists when participants repeatedly search for the same type of target (cups) from the same type of scenes (kitchens) across all trials.

Method

Participants

Twenty Lancaster University students participated in the experiment for course credit. All participants had normal or corrected-to-normal vision and did not participate in any of the previous experiments.

Stimuli

The stimuli consisted of 40 kitchen scenes taken from the Labelme database. Kitchen scenes subtended 24.24° x 18.18° of visual angle when viewed from 57 cm and each scene contained a unique target cup with an average size of 3.42° x 3.46°. Each cup was located somewhere in the scene in a typically occurring location (e.g., on counter surfaces, tables, in open cupboards etc). Otherwise all procedures were identical to Experiment 2 and target cups were never present during previews.

Results

All data were subjected to 2 x 2 repeated measures ANOVA with preview (scene, mask) and target order (before preview, after preview) as within-subjects factors. For all measures, only eye movement metrics associated with correct responses were analysed. A summary of means can be seen in Table 3.

<< Insert Table 3 about here >>

Performance

Search accuracy averaged 80% and did not differ as a function of preview, $F(1, 19) = 1.31$, $p = .267$, $\eta_p^2 = .07$, target order $F(1, 19) = 1.67$, $p = .211$, $\eta_p^2 = 0.04$, and there was no interaction, $F(1, 19) = 0.82$, $p = .376$, $\eta_p^2 = 0.04$. RTs averaged 4784ms across conditions and there was a main effect of preview, $F(1, 19) = 8.03$, $p = .011$, $\eta_p^2 = 0.30$, with participants faster to detect targets in scene preview than mask preview. However there was no main effect of target order, $F(1, 19) = 0.13$, $p = .728$, $\eta_p^2 = 0.01$, and no interaction, $F(1, 19) = 0.21$, $p = .651$, $\eta_p^2 = 0.01$.

Search-related eye movements

Search latency averaged 3934ms across conditions and there was a main effect of preview, $F(1, 19) = 6.12$, $p = .023$, $\eta_p^2 = 0.24$, with participants faster to detect targets in scene preview than mask preview. However there was no main effect of target order, $F(1, 19) = 0.91$, $p = .766$, $\eta_p^2 = .01$, and no interaction, $F(1, 19) = 0.34$, $p = .565$, $\eta_p^2 = 0.02$. The number of fixations averaged 14.42 across conditions and there was a main effect of preview, $F(1, 19) = 8.80$, $p = .008$, $\eta_p^2 = 0.32$, with targets fixated in fewer fixations when participants were presented with a scene preview than a mask preview. There was no main effect of target order, $F(1, 19) = 0.25$, $p = .621$, $\eta_p^2 = 0.01$, and no interaction, $F(1, 19) = 0.56$, $p = .465$, $\eta_p^2 = 0.03$.

First eye movement of search

The initial saccadic latency averaged 241ms across conditions and there was a main effect of preview, $F(1, 19) = 12.57$, $p = .002$, $\eta_p^2 = 0.40$, but no main effect of target order, $F(1, 19) = 0.91$, $p = .351$. $\eta_p^2 = 0.04$, and no interaction, $F(1, 19) = 0.28$, $p = .602$, $\eta_p^2 = 0.01$. The first eye movement of search was executed 32ms quicker when participants were presented with a scene preview than a mask preview, but there was no difference depending on whether the target word was presented before rather than after preview.

The initial saccadic amplitude averaged 1.84° visual angle across conditions and there was a main effect of preview, $F(1, 19) = 7.72$, $p = .012$, $\eta_p{}^2 = 0.29$, but no main effect of target order, $F(1, 19) = 0.02$, $p = .881$, $\eta_p{}^2 = 0.01$. There was however a significant interaction, $F(1, 19) = 11.51$, $p = .003$, $\eta_p{}^2 = 0.38$. When the target word was presented after a preview (i.e., typical FPMW conditions), there was no difference in initial saccadic amplitude between scene preview and mask preview, $F(1, 19) = .02$, $p = .881$, $\eta_p{}^2 < 0.01$, however, there was a difference in initial saccadic amplitude when the target word was presented before a preview, with larger amplitudes for scene compared to mask previews, $F(1, 19) = 29.23$, $p < .001$, $\eta_p{}^2 = 0.61$. Moreover, initial saccadic amplitude in target-scene was larger than scene-target, $F(1, 19) = 5.88$, $p = .025$, $\eta_p{}^2 = 0.24$, and in addition, initial saccadic amplitude in target-mask was also smaller than mask-target, $F(1, 19) = 4.81$, $p = .041$, $\eta_p{}^2 = 0.20$. This suggests that even with repeated search, initial saccadic amplitude was more sensitive to changes in the preview type when the target word was presented before a preview. As to whether these larger initial eye movements were closing in on the direction of the target, distance to target was shortened as a function of preview, $F(1, 19) = 8.06$, $p = .011$, $\eta_p{}^2 = 0.30$, but not target order, $F(1, 19) = 1.43$, $p = .247$. $\eta_p{}^2 = 0.07$, and there was no interaction $F(1, 19) < .01$, $p > .999$, $\eta_p{}^2 < 0.01$. With scene previews the first eye movement was significantly closer to the cup targets compared to mask previews.

Discussion

Experiment 3 provides new evidence to show that the scene preview benefit in the FPMW paradigm is maintained even when repeatedly searching through similar scenes for similar targets. All metrics showed that search was executed more effectively when shown a scene preview. There were no additional benefits in how search was executed for target-scene compared with scene-target conditions, which is consistent with Experiment 2 when targets are removed from preview. Moreover, metrics indicating how search was initiated were also consistent with Experiment 2, in that initial saccadic latencies were quicker with previews, whereas initial saccadic amplitude was still sensitive to our manipulations of target order, even though the target was not present during previews. However, in terms of whether the initial eye movement was closer towards the target, this only seemed to be the case when scene previews were presented.

Experiment 4

One potential confound of the previous experiments is that in addition to manipulating the order that target identity information was presented, there were differences in the integration time available in Scene/Mask-Target and Target-Scene/Mask conditions. As shown in Figure 1, in the Scene-Target FPMW the onset of preview to the onset of search was 2200ms whereas this same interval for Target-Scene FPMW was 700ms. These integration times were chosen based on previous work that tried to accommodate scene perception and medical image perception practices (Litchfield & Donovan, 2016). Related research on FPMW has shown that integration time could be a key factor that underpins search benefits (Võ & Henderson, 2010; Litchfield & Donovan, 2017) and so this discrepancy needs to be resolved. To help isolate the effects of target order from these differences in integration times, a fourth experiment replicated the effects of Experiment 3 whilst also controlling for integration times across all conditions.

Method

Participants

Twenty Edge Hill University students participated in the experiment for course credit. All participants had normal or corrected-to-normal vision and did not participate in any of the previous experiments.

Stimuli

The stimuli consisted of the same 40 kitchen scenes and targets as Experiment 3 and once again the target cups were never present during previews. The key difference in Experiment 4 was that the time from scene preview until onset of search was now held constant across all 4 conditions. To achieve this, for Scene-Target and Mask-Target conditions (typical FPMW) where the target word is presented after preview, participants were presented with a fixation cross for 200ms, and then shown either a scene preview or mask preview for 250ms, followed by a mask for 50ms. A black target word indicated the identity of the target (cup) was then presented for 1300ms, followed by a fixation cross for 200ms, after which windowed search of the scene began. This meant there was a 1800ms delay from the onset of preview, to the onset of search. For Target-Scene and Target-Mask conditions where the target word was presented before preview, participants were presented the target word for 1300ms followed by a fixation cross for 200ms. Immediately afterwards a scene preview or mask preview was presented for 250ms, followed by a mask for 50ms. A final fixation cross was presented for 1500ms, after which windowed search of the scene began. This once again meant there was a 1800ms delay from the onset of preview and the onset of search. It also meant that the target word was onscreen for the same duration across all conditions (1300ms).

Results

All data were subjected to 2 x 2 repeated measures ANOVA with preview (scene, mask) and target order (before preview, after preview) as within-subjects factors. For all measures, only eye movement metrics associated with correct responses were analysed. A summary of means can be seen in Table 4.

<< Insert Table 4 about here >>

Performance

Search accuracy averaged 75% and did not differ as a function of preview, $F(1, 19) = 0.43$, $p = .519$, $\eta_p^2 = .02$, target order, $F(1, 19) = 1.47$, $p = .240$, $\eta_p^2 = .07$, and there was no interaction, $F(1, 19) = 0.29$, $p = .597$, $\eta_p^2 = 0.02$. RTs averaged 4628ms across conditions and there was a main effect of preview, $F(1, 19) = 6.71$, $p = .018$, $\eta_p^2 = 0.26$ with participants faster to detect targets in scene preview than mask preview. However there was no main effect of target order, $F(1, 19) = 0.10$, $p = .759$, $\eta_p^2 = 0.01$, and no interaction, $F(1, 19) = 0.28$, $p = .602$, $\eta_p^2 = 0.02$.

Search-related eye movements

Search latency averaged 3887ms across conditions and there was a main effect of preview, $F(1, 19) = 7.65$, $p = .012$, $\eta_p^2 = 0.29$, with participants faster to detect targets in scene preview than mask preview. However there was no main effect of target order, $F(1, 19) = 0.01$, $p = .940$, $\eta_p^2 < .01$, and no interaction, $F(1, 19) = 1.75$, $p = .201$, $\eta_p^2 = 0.08$. The number of fixations averaged 15.84 across conditions and there was a main effect of preview, $F(1, 19) = 6.73$, $p = .018$, $\eta_p^2 = 0.26$, with targets fixated in fewer fixations when participants were presented with a scene preview than a mask preview. There was no main effect of target order, $F(1, 19) < 0.01$, $p = .984$, $\eta_p^2 < 0.01$, and no interaction, $F(1, 19) = 1.65$, $p = .215$, $\eta_p^2 = 0.08$.

First eye movement of search

The initial saccadic latency averaged 221ms across conditions and there was a main effect of preview, $F(1, 19) = 4.84$, $p = .040$, $\eta_p^2 = 0.20$, and a main effect of target order, $F(1, 19) = 5.42$, $p = .031$. $\eta_p^2 = 0.22$, but no interaction, $F(1, 19) = 0.99$, $p = .332$, $\eta_p^2 = 0.05$. The first eye movement of search was executed 20ms quicker when participants were presented with a scene preview ($M = 210$ms) than a mask preview ($M = 230$ms). Unlike Experiment 3, the first eye movement of search was also executed 27ms faster in the Scene-Target conditions ($M = 207$ms) than the Target-Scene conditions ($M = 234$ms) regardless of preview type.

The initial saccadic amplitude averaged 1.82° visual angle across conditions and there was a main effect of preview, $F(1, 19) = 10.96$, $p = .004$, $\eta_p^2 = 0.37$, with larger amplitudes for scene preview than mask preview. However, there was no main effect of target order, $F(1, 19) = 0.22$, $p = .646$, $\eta_p^2 = 0.01$, and no interaction, $F(1, 19) = 0.17$, $p = .689$, $\eta_p^2 = 0.01$. With regard to whether these initial eye movements were actually closer to the target, distance to cup targets was once again shortened as a function of preview, $F(1, 19) = 16.12$, $p = .001$, $\eta_p^2 = 0.46$, but also target order, $F(1, 19) = 8.75$, $p = .008$, $\eta_p^2 = 0.32$. There was no interaction $F(1, 19) = 1.61$, $p = .221$, $\eta_p^2 = 0.08$. Consistent with Experiment 3 scene previews led to the first eye movement being closer towards the target location, but distance to target was also shortened when the target word was presented before preview.

Discussion

Experiment 4 replicated Experiment 3 but the delay between preview onset and search onset was fixed to 1800ms for all conditions to help ensure differences in integration times did not conflate with target order manipulations. The performance (accuracy, RT) and search metrics (search latency, number of fixations) showed the same pattern of findings as Experiment 3, with a scene preview effect prevailing once again despite the target and scene identity being held constant. This confirms that search benefits of Experiment 3 were not due to differences in integration time and supports recent work showing that scene preview effects can still be found when target identity and scene type is held constant (Litchfield & Donovan, 2016). The only differences between Experiment 3 and 4 related to the first eye movement of search; in Experiment 3 and 4 the first eye movement was faster for scene preview, but in Experiment 4 the first eye movement was also faster in the Scene-Target than the Target-Scene condition. In addition, the amplitudes of the first eye movement in Experiment 3 and 4 were larger in scene preview, but the interaction in Experiment 3 was not found in Experiment 4. This suggests that when integration time was controlled target order still affected the first of eye movement search, and that the interaction identified in Experiment 3 may be due to integration time.

General Discussion

The present study primarily examined different iterations of the FPMW paradigm (Castelhano & Henderson, 2007; Litchfield & Donovan, 2016) to determine how knowing the target identity before or after a brief glimpse of the scene affected initial scene processing and subsequent search. The study also investigated the impact of having target visible during previews, and whether the scene preview benefit still existed when repeatedly searching through the same type of scenes for the same type of object. Consistent with previous studies (e.g., Castelhano & Henderson, 2007; Litchfield &

Donovan, 2016; Võ & Henderson, 2010; Võ & Schneider, 2010) Experiment 1 and 2 found that when compared to mask preview, a scene preview led to reliable improvements in the efficiency and initiation of target search. Moreover, in Experiment 1 we found that knowing the target identity prior to preview led to further improvements in how search was initiated and executed. This provides the first evidence to show that the FPMW paradigm can not only reveal how the initial scene representation can be retained in memory and used to plan subsequent eye movements, but also how this initial scene representation may be influenced by what we expect to see.

It seems that target knowledge can be integrated either prospectively or retrospectively with initial scene information to guide search. Resolving what we see, and what we expect to see, is important in making sense of what we eventually perceive (Summerfield & Egner, 2009; Summerfield, & de Lange, 2014). By integrating target and scene information, fixation placement can be better planned to meet the demands of the task as the search target can be used as a referent for incoming visual processing (Ehinger, Hidalgo-Sotelo, Torralba, & Oliva, 2009). Providing the target identity prior to preview may have activated a prototypical representation of the target object, as well as related semantic knowledge that would shape the expectations of the upcoming scene and the likely spatial associations between objects (Biederman et al., 1982). This prior activation of target knowledge could serve to constrain incoming global information (Castelhano & Henderson, 2007; Wolfe et al., 2011), segment the encoding of the initial scene representation (Zelinksy & Schmidt, 2009), and allow for early target-feature detection (Treisman, 2006). However, since the additional improvements in saccadic search latency in the target-scene condition were not evident in Experiment 2 when the target object was removed from scene previews, prior activation of target knowledge may have primed expectations of where to look and how to encode the scene, but it was early target-feature detection that was the major contributing factor to these improvements.

Knowing the target identity beforehand could have allowed observers to establish a better target template, or more time to process and integrate the target and scene information (Malcolm & Henderson, 2009). Whilst targets were not fixated faster with prior activation of knowledge in Experiment 2, there was a general reaction time advantage for both scene and mask preview conditions when target identity was known beforehand. The fact that this general advantage in reaction time was not evident in Experiment 1 when targets were visible during preview suggests that further studies are necessary to clarify the role of sub-component processes of reaction time. For example, there have been important theoretical distinctions made about verification vs. attentional guidance processes (Castelhano & Heaven 2010; Malcom & Henderson, 2009). Future research could examine these behaviourally defined processes under different FPMW conditions, and also contrast this with how such segmented search is affected when peripheral vision is not restricted (e.g., Hillstrom et al., 2012).

Eye movement metrics relating to the initiation of search offer unique insights that complement the search-related eye movement behaviour (Castelhano, Mack, & Henderson, 2009; Tatler, Baddeley, & Vincent 2006). The initial saccadic latency represents the visual and cognitive processing necessary to decide which potential target location is selected for eye movement and is based on a wealth of information acquired within this time, such as scene-gist and spatial layout (Castelhano & Henderson, 2007; Greene & Oliva, 2009; Tatler et al., 2017). Our finding that initial saccadic latency was consistently reduced following a brief scene preview suggests that such information was acquired and facilitated the planning and initiation of subsequent fixation placement. In addition, participants were quicker to initiate search when the target identity was known beforehand, regardless of whether participants viewed a scene preview or mask preview. This suggests prior activation of target knowledge facilitates the initiation of search towards expected target locations.

The extent to which scene preview and target order led to better planning of fixations or greater readiness to initiate search was also evident in the initial saccadic amplitude. The effect of scene preview on initial saccadic amplitude mirrored the findings of initial saccadic latency, with larger amplitudes accompanied by shorter latencies. However, target order led to increases in initial saccadic amplitude following scene preview, but not mask preview conditions in Experiments 1 and 2. In fact, initial saccadic amplitude was larger in the target-scene condition than the scene-target condition. The combination of greater search efficiency and larger initial saccadic amplitude for the target-scene condition could have implied that all these measures reflected early target-feature search, or even target detection. However, this larger eye movement was still found in Experiment 2 when target detection could not have occurred, and where the additional improvements in search efficiency were negated. We established that the initial eye movement was more likely to be heading towards the target when the target was visible during scene preview, and when participants were aware of the target prior to preview. The fact that this was not the case when the target was absent from scene previews suggests that the larger initial saccadic amplitude for the target-scene condition reflected increased readiness to fixate a potential target site in the periphery. However, in Experiment 3 and 4 when the target was always a cup and never visible during preview, the first eye movement was consistently closer to the target location in scene preview conditions. Experiment 4 also found that having the target word appear first led to closer first fixations relative to target location. Although these initial fixations were still relatively close to the centre of the screen, the fact they were influenced by scene preview and target order supports previous research (e.g., Hillstrom et al., 2012) showing that the initial saccadic amplitude is an acutely sensitive indicator of initial scene processing.

Experiment 3 provided new evidence that the scene preview benefit in FPMW is maintained when participants repeatedly search through the same type of scene for the same type of target, despite the fact that the target was never present during previews. Experiment 4 also confirmed that the search benefits of Experiment 3 were not due to differences in integration time from preview to search. These are key findings as they demonstrate that repeated search does not eliminate the scene preview benefit when using the FPMW paradigm to search for real-world targets. The fact that target order had limited effects on search efficiency across experiments suggests that the order in which knowledge was activated appeared less important than having the right information in the scene preview to exploit (i.e., targets needed to be present during previews to exploit previously activated target knowledge). This sheds new light on recent research that manipulated the contents of target and scene knowledge itself, rather than the order in which target knowledge is activated (Litchfield & Donovan, 2016). If search is guided by target templates (Malcom & Henderson, 2009) and scene-gist information (Bahle, Matsukura, & Hollingworth, in press) then the effectiveness of search should be dependent on the expertise of the individual and their ability to take advantage of this knowledge (Castelhano & Witherspoon, 2016). The findings from the present study show that weak scene preview effects previously observed by Litchfield and Donovan (2016) are unlikely to be due to methodological issues relating to repeated search or knowing beforehand the identity of the upcoming image and target. Instead, what appears to be driving such effects is the difficulty of their medical image perception task in that their targets were both difficult to identify and previewing medical scenes did not sufficiently narrow down the likely locations of these targets. Global processing may be exploited to rapidly categorize an image (Kundel & Nodine, 1975), but this does not mean such processing directly supports localization and recognition of objects (Koehler, & Eckstein, 2017; Litchfield & Donovan, 2017). Experts can make superior detection decisions than novices from brief flashes of medical images, but these same experts are only at chance when subsequently locating the target using blank outlines (Evans, Georgian-Smith, Tambouret, Birdwell, & Wolfe, 2013). In FPMW only correct trials are analysed based on the confirmed localization of the target in search. Prior activation of knowledge may facilitate scene category decisions from rapid presentations (Fabre-Thorpe, 2011), but participants in FPMW are not asked to provide explicit scene categorizations prior to search. Further studies that manipulate expertise of target/scene knowledge are required.

One aspect of our findings that could be due to repeated search is that search metrics in Experiment 3 and 4 were increased relative to Experiment 2. Search latencies and RTs in the former experiments were nearly a second longer, and the preview effect was approximately half of that observed in Experiment 2. This lengthening in search could reflect the impact of repeated search or simply differences between the two testbanks used, with the cups being more difficult to find than the array of targets used in Experiment 1 and 2. Future studies that systematically manipulate target difficulty and type of scene in FPMW conditions may be able to shed further light on this issue.

Many real-world visual search tasks involve repeated search and activation of target templates prior to visual input (Malcolm, Groen, & Baker, 2016), and there has been recent interest in examining these effects of repeated search (Josephs et al., 2014; Võ & Wolfe, 2012; Wolfe & Horowitz, 2017), as well as how this may extend to virtual environments (Kit et al., 2014). Where we should look in a scene to find a target is going to be dictated by our expectations that have been built up by repeated exposure and interaction with particular scenes and objects (Summerfield & de Lange, 2014). In the traditional FPMW paradigm (Castelhano & Henderson, 2007), having target identity made known after preview can reveal how the first glimpse of a scene can be processed alongside subsequent target information to guide search. However, this particular setup deliberately delays at what point target knowledge is activated and how the initial scene may be processed. By not allowing a target template to be activated prior to visual input, this traditional FPMW setup may not at first appear conducive to studying repeated search. In contrast, the findings from the present study show that by adopting a modified sequence of the FPMW whereby target identity is known before scene previews, not only is it possible to investigate repeated search, but that scene preview benefits are still found.

In conclusion, our findings reveal that although we can encode the first glimpse of a scene and then integrate this representation with subsequently activated target knowledge to guide our eye movements, if we already know beforehand what we are looking for then we may process the first glimpse of a scene differently and gain additional advantages in search. However, it seems these advantages are more effectively capitalised if the target we are looking for is actually present in the glimpse in the first place. We extend the scope of the FPMW paradigm by demonstrating that it can also be modified to study repeated visual search whilst controlling for the impact of the initial scene representation.

References

Anderson, N. C., Donk, M., & Meeter, M. (2016). The influence of a scene preview on eye movement behavior in natural scenes. *Psychonomic Bulletin & Review, 23,* 1794-1801.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14*, 143–177.

Bahle, B., Matsukura, M., & Hollingworth, A. (in press). Contrasting gist-based and template-based guidance during real-world visual search. *Journal of Experimental Psychology: Human Perception and Performance*.

Buswell, G. (1935). *How people look at pictures*. University of Chicago Press.

Castelhano, M. S., & Heaven, C. (2010). The relative contribution of scene context and target features to visual search in scenes. *Attention, Perception, & Psychophysics, 72*, 1283–1297.

Castelhano, M. S., & Heaven, C. (2011). Scene context influences without scene gist: Eye movements guided by spatial associations in visual search. *Psychonomic Bulletin & Review, 18*, 890-896.

Castelhano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 753-763.

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision, 9*, 1-15.

Castelhano, M. S., & Witherspoon, R. L. (2016). How you use it matters: Object function guides attention during visual search in scenes. *Psychological Science, 27*, 606-621.

Donovan, T., & Litchfield, D. (2013). Looking for cancer: Expertise related differences in searching and decision making. *Applied Cognitive Psychology, 27*, 43-49.

Drew, T., Evans, K. K., Võ, M. L. H., Jacobson, F. L., & Wolfe, J. M. (2013). What can you see in a single glance and how might this guide visual search in medical images? *Radiographics, 33*, 263–274.

Drew, T., Võ, M. L. H., & Wolfe, J. M. (2013). The Invisible Gorilla Strikes Again Sustained Inattentional Blindness in Expert Observers. *Psychological Science, 24*, 1848-1853.

Ehinger, K., Hidalgo-Sotelo, B., Torralba, A., & Oliva, A. (2009). Modelling search for people in 900 scenes: A combined source model of eye guidance. *Visual Cognition, 17*, 945-978.

Evans, K. K., Georgian-Smith, D., Tambouret, R., Birdwell, R. L., & Wolfe, J. M. (2013). The gist of the abnormal: Above-chance medical decision making in the blink of an eye. *Psychonomic Bulletin & Review, 20*, 1170-1175.

Fabre-Thorpe, M. (2011). The characteristics and limits of rapid visual categorization. *Frontiers in Psychology, 2*, 243.

Greene, M. R., & Oliva, A. (2009). The briefest of glances. The time course of natural scene understanding. *Psychological Science, 40*, 464-472.

Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science, 16*, 219-222.

Hillstrom, A. P., Schloley, H., Liversedge, S. P., Benson, V. (2012). The effect of the first glimpse at a scene on eye movements during search. *Psychomic Bulletin & Review, 19*, 204-210.

Hollingworth, A. (2009). Two forms of scene memory guide visual search: Memory for scene context and memory for the binding of target object to scene location. *Visual Cognition, 17*, 237-291.

Intraub, H. (1980). Presentation rate and the representation of briefly glimpsed pictures in memory. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 1-12.

Josephs, E. L., Dejan Draschkow, Wolfe, J. M., & Võ, M. L.-H. (2016). Gist in time: Scene semantics and structure enhance recall of searched objects. *Acta Psychologica, 169*, 100-108.

Kit, D., Katz, L., Sullivan, B., Snyder, K., Ballard, D., & Hayhoe, M. (2014). Eye movements, visual search and scene memory, in an immersive virtual environment. *PLoS One, 9*(4), e94362.

Koehler, K., & Eckstein, M. P. (2017). Beyond scene gist: Objects guide search more than scene background. *Journal of Experimental Psychology: Human Perception and Performance, 43*, 1177-1193.

Kundel, H. L., & Nodine, C. F. (1975). Interpreting chest radiographs without visual search. *Radiology, 116*, 527–532.

Kundel, H. L., Nodine, C. F., Kunant, E. F., & Weinhaus, S. P. (2007). Holistic component of image perception in mammogram interpretation: Gaze-tracking study. *Radiology, 242*, 396-402.

Litchfield, D. & Donovan, T. (2016). Worth a quick look? Initial scene previews can guide eye movements as a function of domain-specific expertise but can also have unforeseen costs. *Journal of Experimental Psychology: Human Perception and Performance. 42* 982–994.

Litchfield, D. & Donovan, T. (2017). The flash-preview moving window paradigm: Unpacking visual expertise one glimpse at a time. *Frontline Learning Research*, *5*, 80-94.

Malcolm, G. L., Groen, I. I., & Baker, C. I. (2016). Making sense of real-world scenes. *Trends in Cognitive Sciences, 20*, 843-856.

Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision, 9*, 1–13.

Nodine, C. F., & Kundel, H. L. (1987). The cognitive side of visual search in radiology. In J. K., O'Regan, A. Levy-Schoen (Eds). *Eye movements: From physiology to cognition* (pp. 573–582). Amsterdam: Elsevier.

Rayner, K., Smith, T. J., Malcom, G. L., & Henderson, J. M. (2009). Eye movements and visual encoding during scene perception. *Psychological Science, 20*, 6-10.

Russell, B. C., Torralba, A., Murphy, K. P., & Freeman, W. T. (2008). LabelMe: a database and web-based tool for image annotation. *International Journal of Computer Vision, 77*, 157-173.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception, 28*, 1059-1074.

Summerfield, C., & Egner, T (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences, 13*, 403-409.

Summerfield, C., & de Lange, T (2014). Expectation in perceptual decision making: neural and computational mechanisms. *Nature Reviews Neuroscience, 15*, 745-756

Tatler, B. W., Baddeley, R. J., & Vincent, B. T. (2006). The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. *Vision Research, 46*, 1857-1862.

Tatler, B. W., Brockmole, J. R. & Carpenter, R. H. S. (2017). LATEST: A model of saccadic decisions in space and time. *Psychological Review, 124*, 267-300.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review, 113*, 766-786.

Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition, 14*, 411-443.

Võ, M. L. H., & Henderson, J. M. (2010). The time course of initial scene processing for eye movement guidance in natural scene search. *Journal of Vision, 10*, 1-13.

Võ, M. L. H., & Henderson, J. M. (2011). Object-scene inconsistencies do not capture gaze: evidence from the flash-preview moving window paradigm. *Attention, Perception & Psychophysics*, 73, 1742–1753.

Võ, M. L. H., & Schneider, J. M. (2010). A glimpse is not a glimpse: Differential processing of flashed scene previews leads to differential target search benefits. *Visual Cognition, 18*, 171-200.

Võ, M. L. H. & Wolfe, J. M. (2012). When does repeated search in scenes involve memory? Looking at versus looking for objects in scenes. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 23-41.

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, 0058. 1-8.

Wolfe, J. M., Võ, M. L.-H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and non-selective pathways. *Trends in Cognitive Sciences, 15*, 77-84.

Yarbus, A. (1967). *Eye movements and vision*. New York: Plenum Press.

Zelinsky, G. J., & Schmidt, J. (2009). An effect of referential scene constraint on search implies scene segmentation. *Visual Cognition, 17*, 1004-1028.

Table 1.

Observer Performance and Eye Movement Measures of Gaze-Contingent Search in Experiment 1 With Target Present in Preview.

Variable	Scene-Target	Mask-Target	Target-Scene	Target-Mask
Accuracy %				
M	84.25	81.25	85.75	83.50
SE	1.93	1.49	2.06	1.71
RT (ms)				
M	3743	4328	3376	4351
SE	129	132	193	152
Search Latency				
M	2741	3202	2320	3379
SE	122	148	182	167
Number of fixations				
M	10.14	11.74	8.54	12.49
SE	0.40	0.55	0.57	0.56
Initial Saccadic Latency				
M	219	254	194	246
SE	9	10	9	11
Initial Saccadic Amplitude				
M	2.13	1.85	2.60	1.74
SE	0.09	0.10	0.13	0.11

Table 2.

Observer Performance and Eye Movement Measures of Gaze-Contingent Search in Experiment 2 with Target Not Present in Preview.

Variable	Scene-Target	Mask-Target	Target-Scene	Target-Mask
Accuracy %				
M	85.00	83.25	84.75	82.00
SE	1.66	1.67	1.87	2.28
RT (ms)				
M	3760	4514	3369	4442
SE	180	224	176	231
Search Latency				
M	2692	3370	2337	3322
SE	140	216	138	208
Number of fixations				
M	10.01	12.15	8.94	12.16
SE	0.58	0.77	0.49	0.75
Initial Saccadic Latency				
M	227	282	190	268
SE	7	14	7	13
Initial Saccadic Amplitude				
M	2.24	1.97	2.67	1.88
SE	0.09	0.11	0.14	0.10

Table 3.

Observer Performance and Eye Movement Measures of Repeated Gaze-Contingent Search in Experiment 3 With Target Not Present in Preview.

Variable	Scene-Target	Mask-Target	Target-Scene	Target-Mask
Accuracy %				
M	79.50	84.50	77.50	79.00
SE	3.03	2.46	2.89	3.15
RT (ms)				
M	4496	5160	4510	4969
SE	228	214	283	317
Search Latency				
M	3658	4281	3706	4090
SE	235	204	281	294
Number of fixations				
M	13.22	16.06	13.30	15.11
SE	.68	0.77	0.92	1.06
Initial Saccadic Latency				
M	224	250	226	264
SE	12	11	13	13
Initial Saccadic Amplitude				
M	1.86	1.83	2.18	1.48
SE	0.10	0.18	0.10	0.07

Table 4.

Observer Performance and Eye Movement Measures of Repeated Gaze-Contingent Search in Experiment 4 With Target Not Present in Preview.

Variable	Scene-Target	Mask-Target	Target-Scene	Target-Mask
Accuracy %				
M	73.50	73.00	78.00	74.50
SE	3.72	3.33	2.87	2.85
RT (ms)				
M	4285	5047	4317	4866
SE	307	265	364	293
Search Latency				
M	3402	4356	3707	4085
SE	295	262	355	300
Number of fixations				
M	14.09	17.58	15.08	16.62
SE	1.09	0.89	1.32	0.97
Initial Saccadic Latency				
M	204	210	218	251
SE	13	12	12	13
Initial Saccadic Amplitude				
M	2.06	1.51	2.09	1.61
SE	0.16	0.13	0.23	0.10

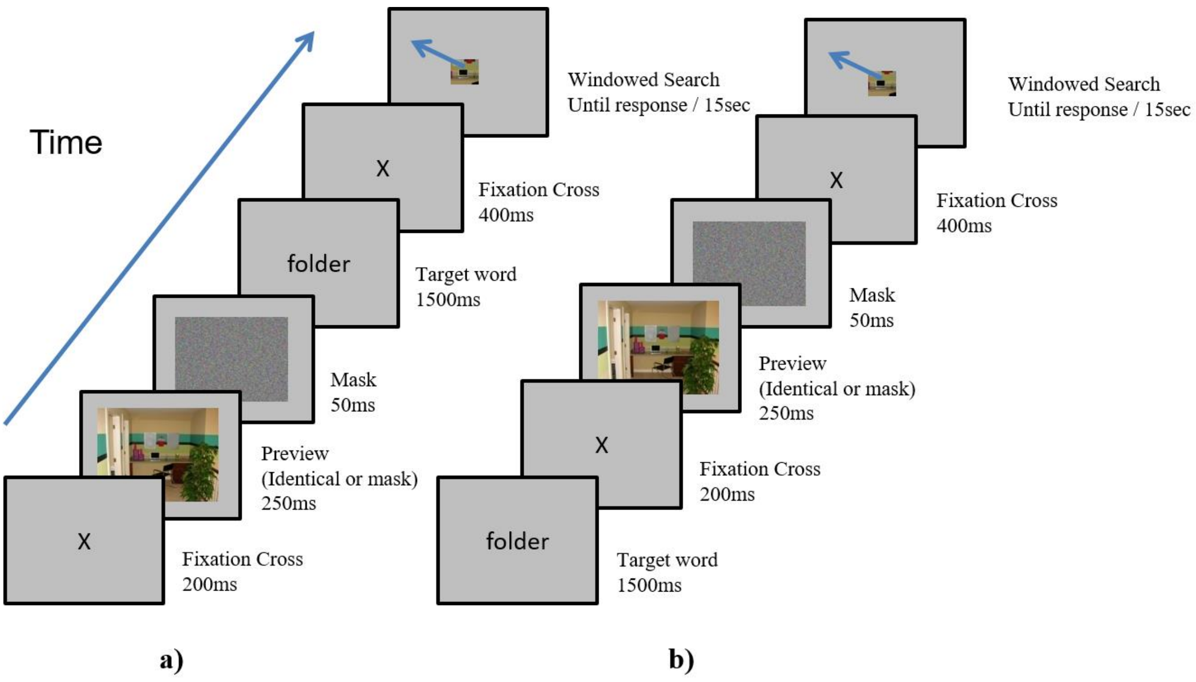

Figure 1. Trial sequence of the ‘flash-preview moving window’ paradigm used in Experiments 1-3 based on whether a) target identity presented after preview, or b) target identity presented before preview.